L5 ANSWER 1 OF 78 CAPLUS COPYRIGHT 2003 ACS
Full Text
AN 1990:4003 CAPLUS
DN 112:4003
TI An **algorithm** for multiple **alignment** of protein **sequences**
AU Park, Kiejung; Sheen, Joonho; Park, Chankyu
CS Dep. Biol. Sci. Eng., Korea Adv. Inst. Sci. Technol., Seoul, 150, S. Korea
SO Han'guk Saenghwa Hakhoechi (1989), 22(3), 346.54
CODEN: KBCJAK; ISSN: 0368.4881
DT Journal
LA Korean
AB One application of **computers** in mol. biol. has been for comparing a large no. of mol. **sequences** in very short time. For this purpose, a new **algorithm** is proposed which differs in several aspects from other approaches. This **algorithm**, called MAlign, is designed to seek global homol. by introducing an effective way to make simultaneous comparisons among test **sequences**. One problem in previous **algorithms** which were limited in its ability to compare **sequences** simultaneously has been solved by introducing intermediate consensus or compacted **sequences** and including them for comparison. In addn., a homol. **vector** concept was applied to provide uniform representation for each intermediate, which makes global comparison easier. Several test results indicate that high homol. values obtained from pairwise **alignment** are maintained after multiple **alignment** of those **sequences**, which is more apparent in higher homol. values. Sample **alignment** results using this approach for three different copper.binding proteins as well as bacterial signaling proteins are presented.

L5 ANSWER 4 OF 78 CAPLUS COPYRIGHT 2003 ACS
Full Text
AN 1995:228028 CAPLUS
DN 122:27196
TI Multiple protein structure **alignment**
AU Taylor, William R.; Flores, Tomas P.; Orengo, Christine A.
CS Lab. Mathematical Biol., Natl. Inst. Med. Res., London, NW7 1AA, UK
SO Protein Science (1994), 3(10), 1858.70
CODEN: PRCIEI; ISSN: 0961.8368
PB Cambridge University Press
DT Journal
LA English
AB A method was developed to compare protein, e.g., IgG, structures and to combine them into a multiple structure consensus. Previous methods of multiple structure comparison have only concatenated pairwise **alignments** or produced a consensus structure by averaging coordinate sets. The current method is a fusion of the fast structure comparison program SSAP and the multiple **sequence alignment** program MULTAL. As in MULTAL, structures are progressively combined, producing intermediate consensus structures that are compared directly to each other and all remaining single structures. This leads to a hierarchic "condensation", continually evaluated in the light of the emerging conserved core regions. Following the SSAP approach, all interat. **vectors** were retained with well.conserved regions distinguished by coherent **vector** bundles (the structural equiv. of a conserved **sequence** position). Each bundle of **vectors** is summarized by a resultant, whereas **vector** coherence is captured in an error term, which is the only distinction between conserved and variable positions. Resultant **vectors** are used directly in the comparison, which is weighted by their error values, giving greater importance to the matching of conserved positions. The resultant **vectors** and their errors can also be used directly in mol. modeling. Applications of the method were assessed by the quality of the resulting **sequence alignments**, phylogenetic tree construction, and data bank scanning with the consensus. Visual assessment of the structural superpositions and consensus structure for various well.characterized families confirmed that the consensus had identified a reasonable core.

L5 ANSWER 14 OF 78 JICST.EPlus COPYRIGHT 2003 JST
Full Text
AN 940876846 JICST.EPlus
TI Substructure Search and **Alignment Algorithms** for Three.Dimensional Protein Structures.
AU AKUTSU T
CS Gunma Univ., Gunma, JPN
SO Joho Shori Gakkai Kenkyu Hokoku, (1994) vol. 94, no. 82(AL.41), pp. 1.8.
Journal Code: Z0031B (Fig. 1, Tbl. 2, Ref. 22)
ISSN: 0919.6072
CY Japan
DT Journal; Article
LA English
STA New
AB This paper presents two practical **algorithms** for pattern matching of 3D protein structures: a hashing technique for quick substructure search and an **alignment algorithm** for 3D structures. In both **algorithms**, protein structures are treated as point **sequences**. In the hashing technique, for each fixed.length **sequence**, a hash **vector** is**computed**, where the distance between two hash **vectors** is small if two **sequences** are similar. In the **alignment algorithm**, a correspondence of points between two **sequences** is **computed**. In each **algorithm**, a theoretical proof for the quality of outputs is given. Moreover, experimental results show that both **algorithms** are effective. (author abst.)

L5 ANSWER 23 OF 78 MEDLINE
Full Text
AN 91332920 MEDLINE
DN 91332920 PubMed ID: 1908023
TI Average values of a dissimilarity measure not requiring **sequence alignment** are twice the averages of conventional mismatch counts requiring **sequence alignment** for a variety of **computer**.generated model systems.
AU Blaisdell B E
CS Department of Mathematics, Stanford University, CA 94305.
SO JOURNAL OF MOLECULAR EVOLUTION, (1991 Jun) 32 (6) 521.8.
Journal code: 0360051. ISSN: 0022.2844.
CY United States
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 199109ED Entered STN: 19911006
Last Updated on STN: 19911006
Entered Medline: 19910917
AB A measure of **sequence** similarity, dt, not requiring prior **sequence alignment** gave correct results for a variety of **computer**.generated model **sequences** without and with gaps for all degrees of substitution, s. Measure d was the squared Euclidean distance between **vectors** of counts of t.tuplets of characters in the two **sequences**. In models without gaps and without Needleman.Wunsch **alignment**, average d was very closely equal to twice average conventional mismatch counts, m. In these models one of each of the conditions on the Jukes.Cantor model was violated in turn: (1) both descendant lineages receive the same number of substitutions, (2) all sites are equally likely to be substituted, (3) all different replacement characters are equally likely to be chosen, and (4) all original characters are equally likely to be substituted. In Jukes.Cantor models with gaps Needleman.Wunsch **alignment** was necessarily performed, a procedure that generally produced incorrect values of m. For these models average d was found to be very closely equal to twice the average m estimated from the known value of s using the inverted Jukes.Cantor formula.

L5 ANSWER 32 OF 78 PASCAL COPYRIGHT 2003 INIST.CNRS. ALL RIGHTS
Full Text
RESERVED.
AN 2003.0168454 PASCAL

TIEN Application of Max.plus algebra to biological **sequence** comparisons
Max.plus algebras
AU COMET J..P.
GAUBERT Stephane (ed.); LOISEAU Jean.Jacques (ed.); MAIRESSE Jean (ed.);
NIVAT Maurice (ed.); PIN Jean.Eric (ed.)
CS LAMI, CNRS UMR 8042, University of Evry, Tour Evry 2, 523 Place des terrasses de l'agora, 91025 Evry, France
INRIA, Rocquencourt, France; CNRS, IRCCyN, Nantes, France; CNRS, LIAFA, Paris, France; Universite Paris 7, LIAFA, Paris, France
SO Theoretical computer science, (2003), 293(1), 189.217, 28 refs.
ISSN: 0304.3975 CODEN: TCSCDI
DT Journal
BL Analytic
CY Netherlands
LA English
AV INIST.17243, 354000107253650100
AB The classical **algorithms** to **align** two biological **sequences** (Needleman and Wunsch and Smith and Waterman **algorithms**) can be seen as a **sequence** of elementary operations in (max, +) algebra: each line (viewed as a **vector**) of the dynamic programming table of the **alignment algorithms** can be deduced by a (max, +) multiplication of the previous line by a matrix. Taking into account the properties of these matrices there are only a finite number of nonproportional **vectors**. The use of this algebra allows one to imagine a faster equivalent **algorithm**. One can construct an automaton and afterwards skim through the **sequence** databank with this automaton in linear time. Unfortunately, the size of the automaton prevents using this approach for comparing global proteins. However, biologists frequently face the problem of comparing one short string against many others **sequences**. In that case this automaton version of dynamic programming results in a new **algorithm** which works faster than the classical **algorithm**.

L5 ANSWER 35 OF 78 MEDLINE
Full Text
AN 2002331987 MEDLINE
DN 22069937 PubMed ID: 12075023
TI SST: an **algorithm** for finding near.exact **sequence** matches in time proportional to the logarithm of the database size.
AU Giladi Eldar; Walker Michael G; Wang James Z; Volkmuth Wayne
CS Incyte Pharmaceuticals, 3174 Porter Drive, Palo Alto, CA 94304, USA..
egiladi@incyte.com
SO BIOINFORMATICS, (2002 Jun) 18 (6) 873.7.
Journal code: 9808944. ISSN: 1367.4803.
CY England: United Kingdom
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200301
ED Entered STN: 20020621
Last Updated on STN: 20030128
Entered Medline: 20030127
AB MOTIVATION: Searches for near exact **sequence** matches are performed frequently in large.scale sequencing projects and in comparative genomics. The time and cost of performing these large.scale **sequence**.similarity searches is prohibitive using even the fastest of the extant **algorithms**. Faster **algorithms** are desired. RESULTS: We have developed an **algorithm**, called SST (**Sequence** Search Tree), that searches a database of DNA **sequences** for near.exact matches, in time proportional to thelogarithm of the database size n. In SST, we partition each **sequence** into fragments of fixed length called 'windows' using multiple offsets. Each window is mapped into a **vector** of dimension 4(k) which contains the frequency of occurrence of its component k.tuples, with k a parameter typically in the range 4.6. Then we create a tree.structured index of the windows in **vector** space, with tree.structured **vector** quantization (TSVQ). We identify the nearest neighbors of a query **sequence** by partitioning the query into windows and searching the tree.structured index for nearest.neighbor windows in the database. When the tree is balanced this yields an O(logn) complexity for the search. This complexity was observed in our **computations**. SST is most effective for applications in which the target **sequences** show a high degree of similarity to the query **sequence**, such as assembling shotgun **sequences** or matching ESTs to genomic **sequence**. The **algorithm** is also an effective filtration method. Specifically, it can be used as a preprocessing step for other search methods to reduce the complexity of searching one large database against another. For the problem of identifying overlapping fragments in the assembly of 120 000 fragments from a 1.5 megabase genomic **sequence**, SST is 15 times faster than BLAST when we consider both building and searching the tree. For searching alone (i.e. after building the tree index), SST 27 times faster than BLAST. AVAILABILITY: Request from the authors.

L5 ANSWER 40 OF 78 PASCAL COPYRIGHT 2003 INIST.CNRS. ALL RIGHTS
Full Text
RESERVED.
AN 1999.0514804 PASCAL

TIEN Markovian structures in biological **sequence alignments**
AU LIU J. S.; NEUWALD A. F.; LAWRENCE C. E.
CS Department of Statistics, Stanford University, Stanford, CA 94305, United States; Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, United States; Biometrics Lab, Wadsworth Center for Laboratories and Research, New York State Department of Health, Albany, NY 12201, United States
SO Journal of the American Statistical Association, (1999), 94(445), 1.15, 44 refs.
ISSN: 0162.1459 CODEN: JSTNAL
DT Journal
BL Analytic
CY United States
LA English
AV INIST.3094, 354000083478830010
AB The **alignment** of multiple homologous biopolymer **sequences** is crucial in research on protein modeling and engineering, molecular evolution, and prediction in terms of both gene function and gene product structure. In this article we provide a coherent view of the two recent models used for multiple **sequence alignment**.the hidden Markov model (HMM) and the block.based motif model.to develop a set of new **algorithms** that have both the sensitivity of the block.based model and the flexibility of the HMM. In particular, we decompose the standard HMM into two components: the insertion component, which is captured by the so.called "propagation model," and the deletion component, which is described by a deletion **vector**. Such a decomposition serves as a basis for rational compromise between biological specificity and model flexibility. Furthermore, we introduce a Bayesian model selection criterion that.in combination with the propagation model, genetic **algorithm**, and other **computational** aspects.forms the core of PROBE, a multiple **alignment** and database search methodology. The application of our method to a GTPase family of protein **sequences** yields an **alignment** that is confirmed by comparison with known tertiary structures.

L5 ANSWER 45 OF 78 MEDLINE
Full Text
AN 2002064973 MEDLINE
DN 21650566 PubMed ID: 11791227
TI Local multiple **alignment** of numerical **sequences**: detection of subtle motifs from protein **sequences** and structures.
AU Akutsu T; Horimoto K
CS Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611.0011, Japan.. takutsu@kuicr.kyoto.u.ac.jp
SO GENOME INFORMATICS SERIES, (2001) 12 83.92.
Journal code: 9717234. ISSN: 0919.9454.
CY Japan
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200208
ED Entered STN: 20020125
Last Updated on STN: 20020820
Entered Medline: 20020819
AB This paper presents a new method to find motifs from multiple protein **sequences** and multiple protein structures. The method consists of two parts: quantification and local multiple **alignment**. In the former part, protein **sequences** and protein structures are transformed into **sequences** of real numbers and real **vectors** respectively. In the latter part, fixed length regions having similar shapes are located. A Gibbs sampling **algorithm** for **sequences** of real numbers/**vectors** is newly developed for finding common regions. The results of the comparison with a standard Gibbs sampling program show that the method is particularly useful when structural information is available.

L5 ANSWER 50 OF 78 MEDLINE
Full Text
AN 2002105924 MEDLINE
DN 21825953 PubMed ID: 11836217
TI Integrated gene and species phylogenies from unaligned whole genome protein **sequences**.
AU Stuart Gary W; Moffett Karen; Baker Steve
CS Department of Life Sciences, Indiana State University, Terre Haute, IN 47809, USA.. G.Stuart@indstate.edu
SO BIOINFORMATICS, (2002 Jan) 18 (1) 100.8.
Journal code: 9808944. ISSN: 1367.4803.
CY England: United Kingdom .
DT Journal; Article; (JOURNAL ARTICLE)
LA English
FS Priority Journals
EM 200206
ED Entered STN: 20020212
Last Updated on STN: 20020611
Entered Medline: 20020610
AB MOTIVATION: Most molecular phylogenies are based on **sequence alignments**. Consequently, they fail to account for modes of **sequence** evolution that involve frequent insertions or deletions. Here we present a method for generating accurate gene and species phylogenies from whole genome **sequence** that makes use of short character string matches not placed within explicit **alignments**. In this work, the singular value decomposition of a sparse tetrapeptide frequency matrix is used to represent the proteins of organisms uniquely and precisely as **vectors** in a high.dimensional space. **Vectors** of this kind can be used to calculate pairwise distance values based on the angle separating the **vectors**, and the resulting distance values can be used to generate phylogenetic trees. Protein trees so derived can be examined directly for homologous **sequences**. Alternatively, **vectors** defining each of the proteins within an organism can be summed to provide a **vector** representation of the organism, which is then used to generate species trees. RESULTS: Using a large mitochondrial genome dataset, we have produced species trees that are largely in agreement with previously published trees based on the analysis of identical datasets using different methods. These trees also agree well with currently accepted phylogenetic theory. In principle, our method could be used to compare much larger bacterial or nuclear genomes in full molecular detail, ultimately allowing accurate gene and species relationships to be derived from a comprehensive comparison of complete genomes. In contrast to phylogenetic methods based on **alignments**, **sequences** that evolve by relative insertion or deletion would tend to remain recognizably similar.